

CLEARANCE REQUEST FOR PUBLIC RELEASE OF DEPARTMENT OF DEFENSE INFORMATION

(See Instructions on back.)

(This form is to be used in requesting review and clearance of DoD information proposed for public release in accordance with DoDD 5230.9.)

TO: Director, Freedom of Information & Security Review, Rm. 2C757, Pentagon

1. DOCUMENT DESCRIPTION

a. TYPE
Presentation for Posting to Website

b. TITLE
Uses, Adv, and Disadv of GANs: The Coriolis Case Study

c. PAGE COUNT
53

d. SUBJECT AREA
Cost Analysis

2. AUTHOR/SPEAKER

a. NAME (Last, First, Middle Initial)
Whitaker, York W.

b. RANK
CIV

c. TITLE
Research Fellow

d. OFFICE
NA

e. AGENCY
LMI

3. PRESENTATION/PUBLICATION DATA (Date, Place, Event)

Request permission to post the presentation to the DoD Cost Analyst Symposium website at <http://www.dodcas.osd.mil>. The purpose is to make symposium presentations available to interested parties. The symposium was held on February 14-18, 2005 at the Williamsburg Marriott in Williamsburg, VA.

4. POINT OF CONTACT

a. NAME (Last, First, Middle Initial)
Angers, Jeffrey P.

b. TELEPHONE NO. (Include Area Code)
(703) 692-8045

5. PRIOR COORDINATION

a. NAME (Last, First, Middle Initial)

b. OFFICE/AGENCY

c. TELEPHONE NO. (Include Area Code)

6. REMARKS

CLEARED
For Open Publication

APR 18 2005 3

Office of Freedom of Information
and Security Review
Department of Defense

7. RECOMMENDATION OF SUBMITTING OFFICE/AGENCY

a. THE ATTACHED MATERIAL HAS DEPARTMENT/OFFICE/AGENCY APPROVAL FOR PUBLIC RELEASE (qualifications, if any, are indicated in Remarks section) AND CLEARANCE FOR OPEN PUBLICATION IS RECOMMENDED UNDER PROVISIONS OF DODD 5320.9. I AM AUTHORIZED TO MAKE THIS RECOMMENDATION FOR RELEASE ON BEHALF OF:

Chairman, Cost Analysis Improvement Group

b. CLEARANCE IS REQUESTED BY 20050431 (YYYYMMDD).

c. NAME (Last, First, Middle Initial)
Vogel, Russell A.

d. TITLE
CAIG Executive Secretary

e. OFFICE
OSD PA&E

f. AGENCY
DoD

g. SIGNATURE

Russell A. Vogel

h. DATE SIGNED (YYYYMMDD)

20050405

05-S-1265



GOVERNMENT CONSULTING

THE OPPORTUNITY TO MAKE A DIFFERENCE HAS NEVER BEEN GREATER

USES, ADVANTAGES, AND DISADVANTAGES OF GENERALIZED ACTIVITY NETWORKS



38th Annual DoD Cost Analysis Symposium
Williamsburg, VA
15-18 February 2005

Agenda

- Background
- What is a Generalized Activity Network (GAN)?
- Advantages & Disadvantages to GAN Approach
- GAN Performance vs. SERs: Coriolis Case Study
- Conclusion



Agenda

- **Background**
- What is a Generalized Activity Network (GAN)?
- Advantages & Disadvantages to GAN Approach
- GAN Performance vs. SERs: Coriolis Case Study
- Conclusion



Background

- History: 5+ years LMI and OSD/PA&E research project
- Objective: Improve development program cost estimates
 - Identify and investigate new methods for cost estimation
 - Build tools and models that implement these methods
- Motivation: Development program cost estimation is notoriously difficult



Background

- Generalized Activity Network (GAN) models increasingly used by cost analysts
- Recent Applications
 - Air Traffic Management (ATM) system development
 - Spiral Development Programs
 - Missile Defense Systems
 - Satellite Development Schedules
 - Army Weapon System Test and Evaluation Costs
- Growing use of GAN models necessitates evaluation
 - Performance vs. traditional regression methods
 - Appropriate Use of GAN models



Agenda

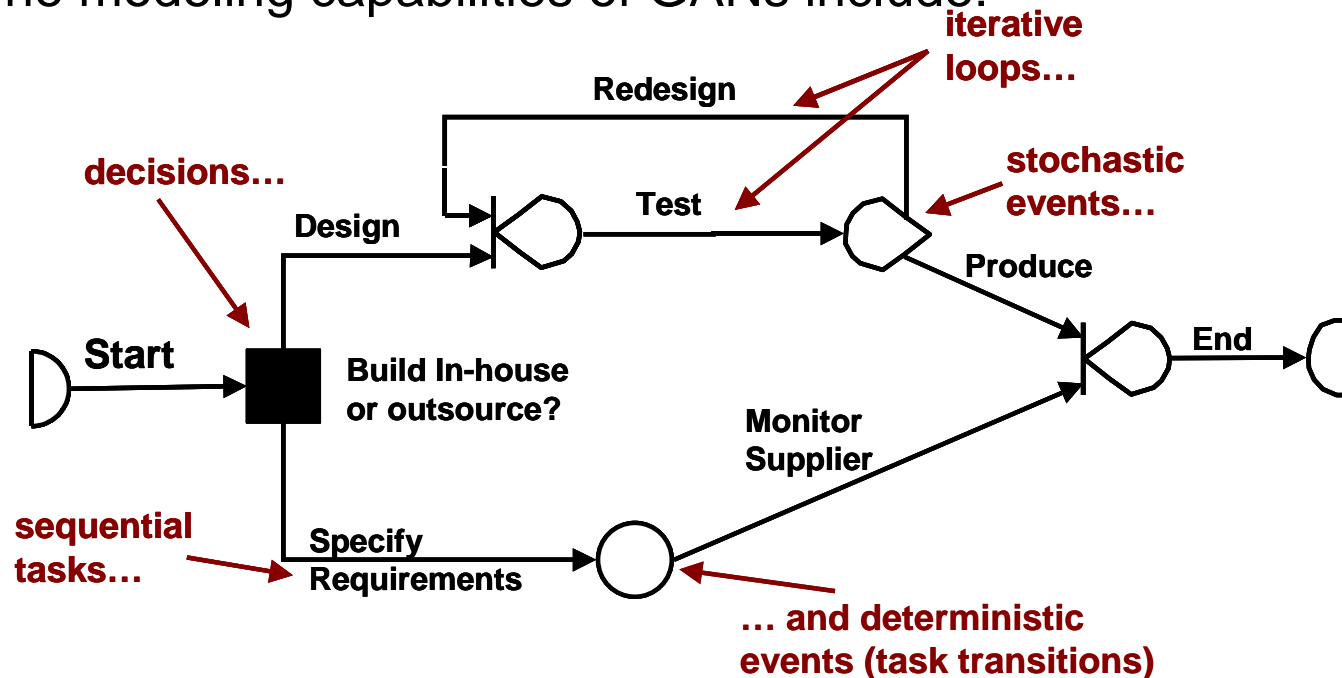
- Background
- What is a Generalized Activity Network (GAN)?
- Advantages & Disadvantages to GAN Approach
- GAN Performance vs. SERs: Coriolis Case Study
- Conclusion



What is a GAN?

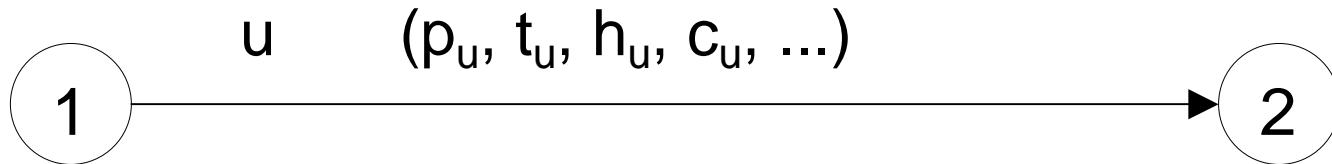
A Generalized Activity Network (GAN) is...

- A cyclical directed process modeling diagram (an extension of PERT)
- The modeling capabilities of GANs include:



What is a GAN?

A GAN has as its basic element an activity (u)



$p_u \equiv$ probability that arc “u” executes

$t_u \equiv$ u’s execution time

$h_u(t_u) \equiv$ probability density function for t

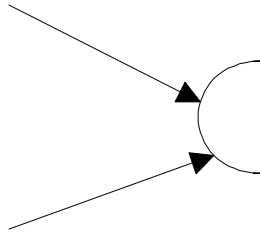
$c_u \equiv$ u’s cost: may depend upon t



GAN Junctions

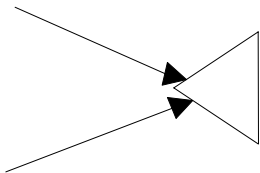
GAN Receivers

**And
(AND)**



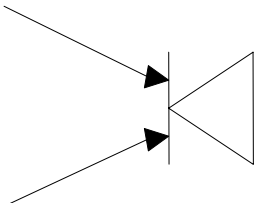
All arcs must
execute to
continue

**Inclusive Or
(OR)**



Continue
after any arc
completes

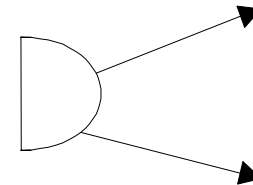
**Exclusive Or
(XOR)**



Must complete
exactly one
arc to continue

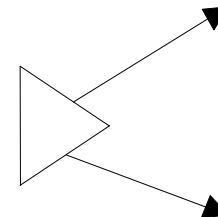
GAN Transmitters

Must follow



All arcs
execute

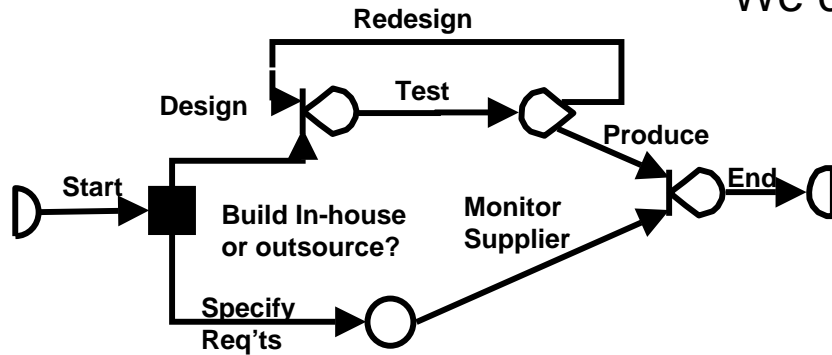
May follow



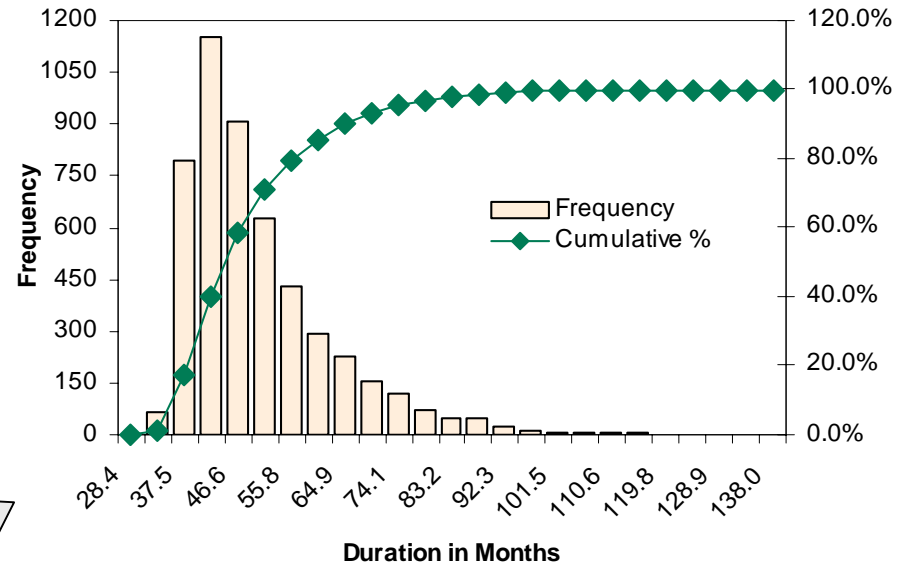
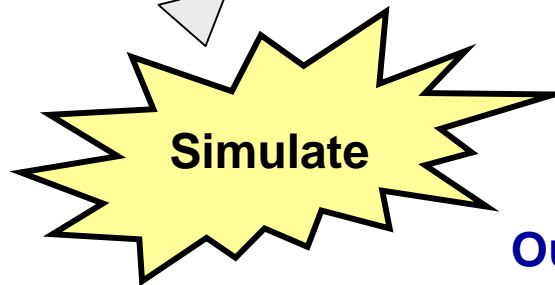
Arcs execute
with assigned
probabilities

GAN-Based Simulations

We convert GANs...



...into simulations...



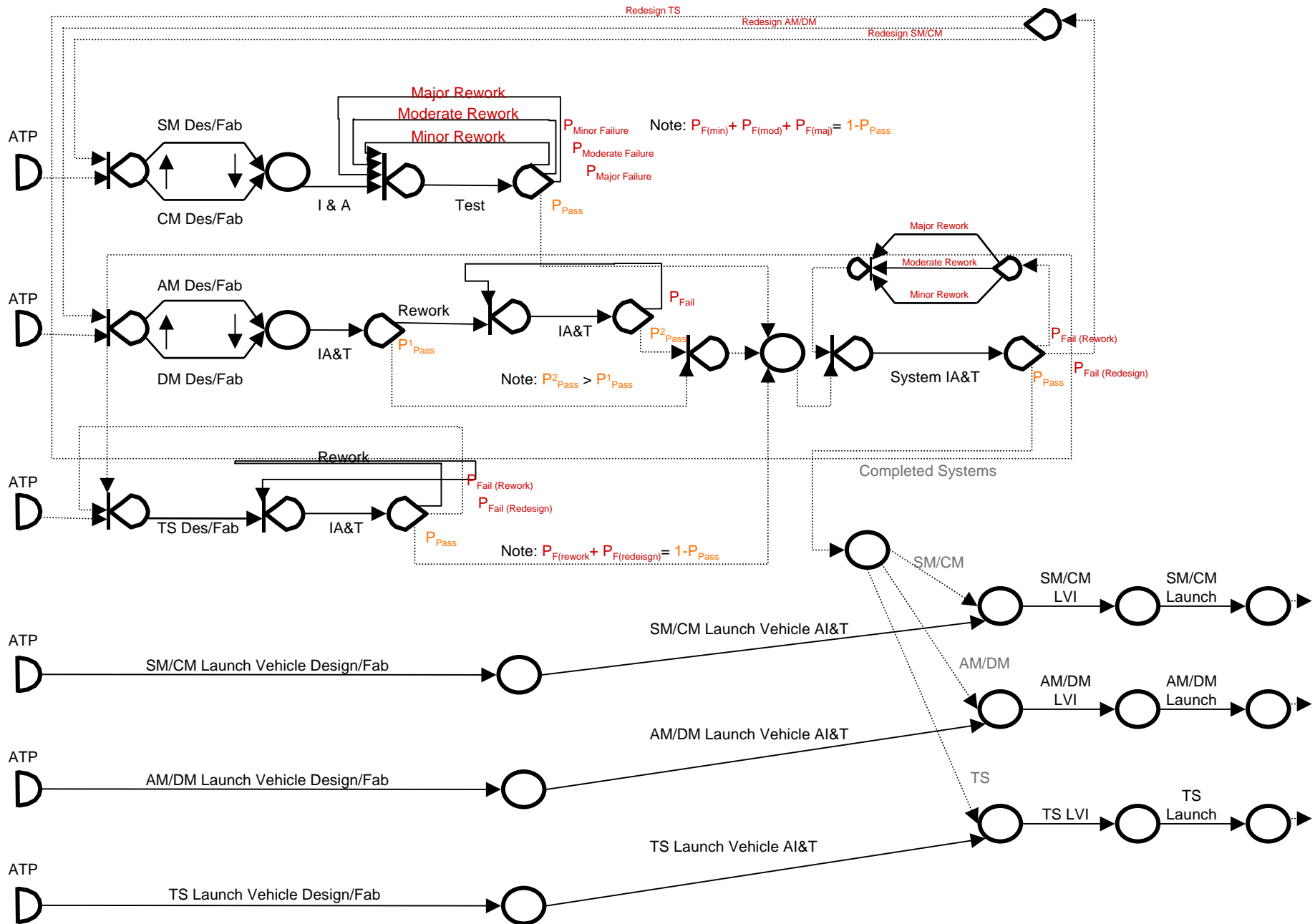
...that can compute completion time and cost for complex spending programs.

Our research shows that these simulations provide a surprising amount of insight, even with few inputs

How GANS are Built and Calibrated

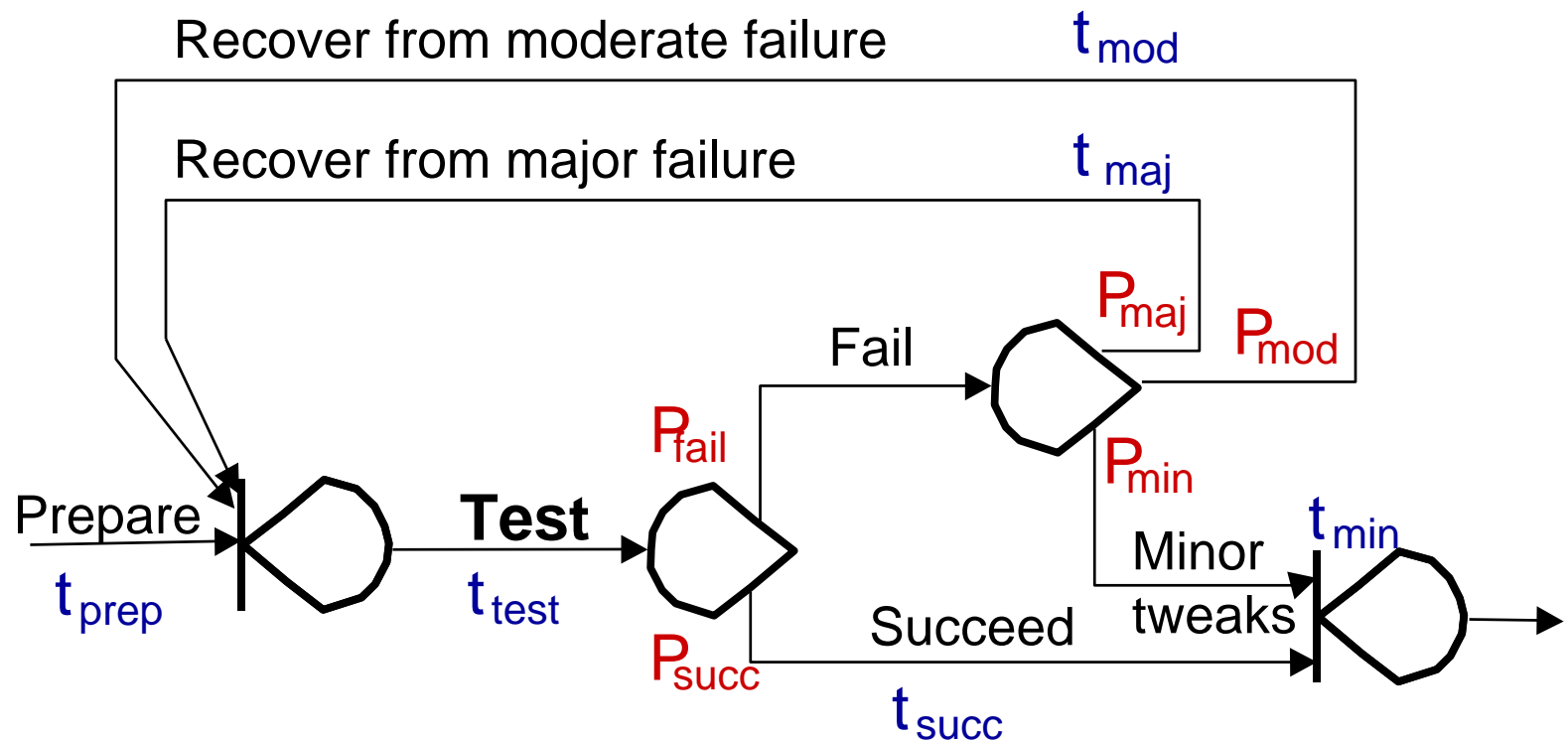
- Modeling process:
 - Build a network diagram (GAN) to describe possible program execution paths
 - Estimate parameters: need random distributions for task durations
 - Require probabilities for feedback loops or other event outcomes
 - Create a discrete-event simulation for that network
- Parameter estimation:
 - Task durations can be based off:
 - Build-up estimates, calibration with historical data, or engineering judgment
 - Usually apply a Weibull distribution (Gladstone-Miller 2002) to deterministic estimate
 - Feedback probabilities can be calibrated with historical data from similar programs or engineering judgment





Example Application

Repeat-Until-Pass Test GAN



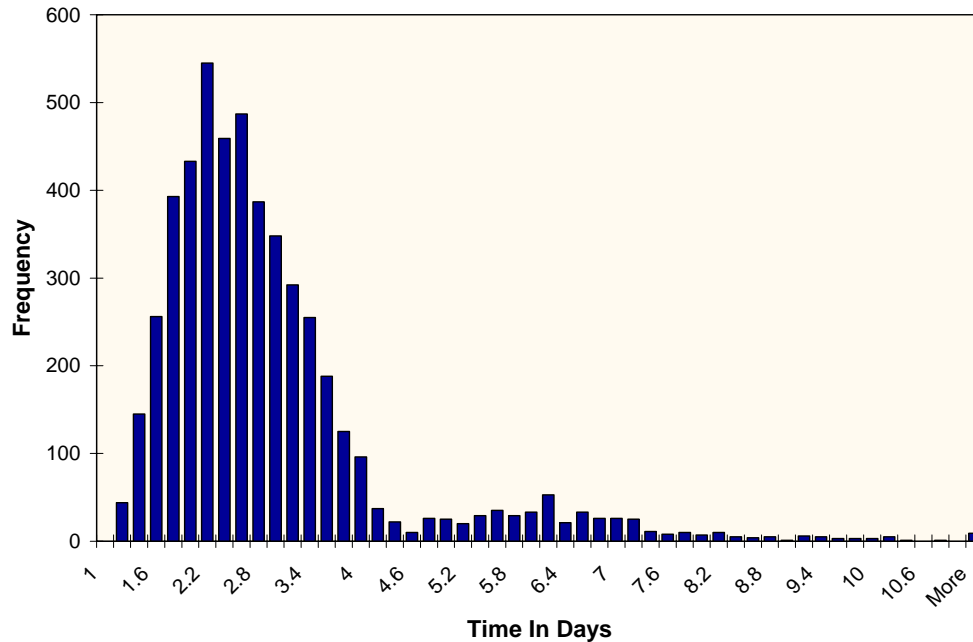
Most likely task durations in blue Event outcome probabilities in red

Example: Repeat-Until-Pass Test GAN

- Durations for Preparation and Testing:
 - Uniform Random Variables
 - Expectation 1 day & Range 1 day: $U(0.5, 1.5)$
- Durations for Recovery from Test Failure:
 - Minor Failure: $U(0.5, 1.5)$ Exp. Value: 1 day
 - Moderate Failure: $U(1.25, 2.75)$ Exp. Value: 2 days
 - Major Failure: $U(2.0, 4.0)$ Exp. Value: 3 days
 - Note: Dispersion also increases with failure severity
- Duration for activities following success is 0
- $P_{\text{success}} = P_{\text{failure}} = .5$
- $P_{\text{min}} = .8$; $P_{\text{mod}} = P_{\text{maj}} = .1$
 - 10% of all failures are moderate, and 10% of all failures are major



Example: Repeat-Until-Pass Test GAN



- Performed Monte Carlo Simulation (5000 Draws)
- Expected Duration for Test Success: 2.8 days
- Large right-tail dispersion due to geometric distribution from inclusion of a probability of test failure



Agenda

- Background
- What is a Generalized Activity Network (GAN)?
- **Advantages & Disadvantages to GAN Approach**
- GAN Performance vs. SERs: Coriolis Case Study
- Conclusion



GAN Advantages

- Hierarchical
- Flexible
- Model iterative processes
- Can provide more information than simple time/cost estimates
 - Complete distribution; eliminates need for separate risk analysis
 - Identify potential problem activities for risk mitigation
- Often provide useful insight during both design (diagramming) and analysis (simulation, analytic equations) phases
- Often “force” analyst to consider program/process from more detailed perspective



GAN Advantages

- Discipline of creating one helps identify/clarify critical issues up front
- Can calibrate model to estimate
 - Time at completion
 - Cost at completion
 - Quality of product at completion
- Shows how activities interact (through GAN junctions)
- Takes mystery away from integration processes



GAN Disadvantages

- Requires large amount of detailed program data
 - Data necessary for calibration
 - Calibration necessary for meaningful cost/schedule estimates
- “Uniqueness” problem
 - Data cannot be used for calibration if too program-specific
 - Breadth of data as important as depth of data
- May suffer from subjectivity of expert opinion data
 - Problem of all bottom-up estimates
- “Familiarity” problem: Although growing, GANs currently not widely used for cost analysis



Agenda

- Background
- What is a Generalized Activity Network (GAN)?
- Advantages & Disadvantages to GAN Approach
- **GAN Performance vs. SERs: Coriolis Case Study**
- Conclusion



GAN vs. SER Performance

- Increasing use of GANs for schedule estimation
- GANs bridge gap between bottom-up and parametric methods
 - Bottom-up methods
 - Detailed but doesn't consider risk and uncertainty
 - Requires extensive data
 - Almost always low
 - Parametric methods: Schedule Estimation Relationships (SERs)
 - Requires little data
 - Provides no managerial information beyond estimate
- GANs may provide additional information but what about predictive performance?



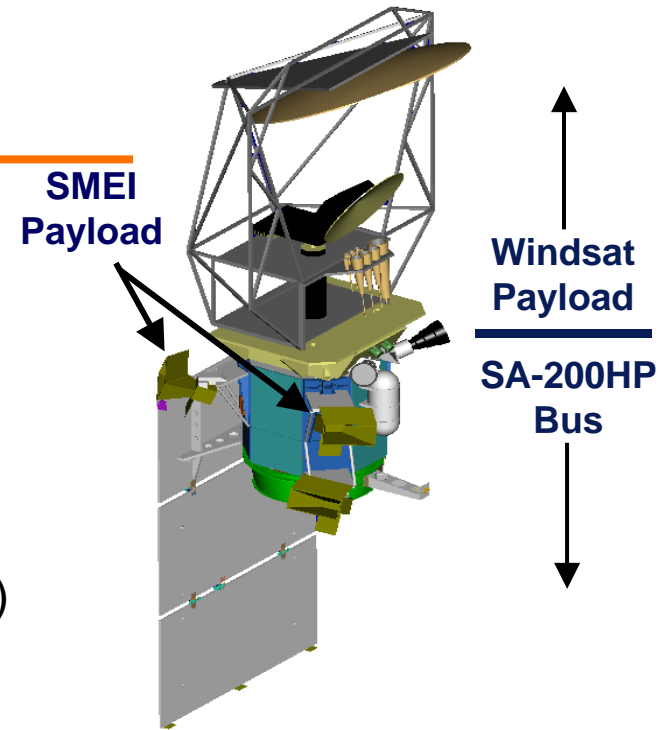
GAN vs. SER Performance: Method

- Tasked to compare relative performance of GANs & SERs
- Identify system for case study
 - Fielded system
 - Data on initial schedule at contract award
- Obtain published commodity specific SERs
- Construct GAN model for system development from initial schedule
- Predict schedule using each approach using data from contract award
- Compare to actual schedule at completion



Case Study: Coriolis Satellite

- Joint Air Force & Navy Development
- Two independent payloads
 - Air Force: Solar Mass Ejection Imager (SMEI)
 - Navy: Windsat
 - Scientific/Sensor Mission
 - Proof of concept for future National Polar-orbiting Operational Environmental Satellite System (NPOESS) missions
- COTS/Heritage Bus: SpectrumAstro SA200-HP
- Initial schedule data reconstructed at ATP
- Successfully launched – can compare predicted to actual

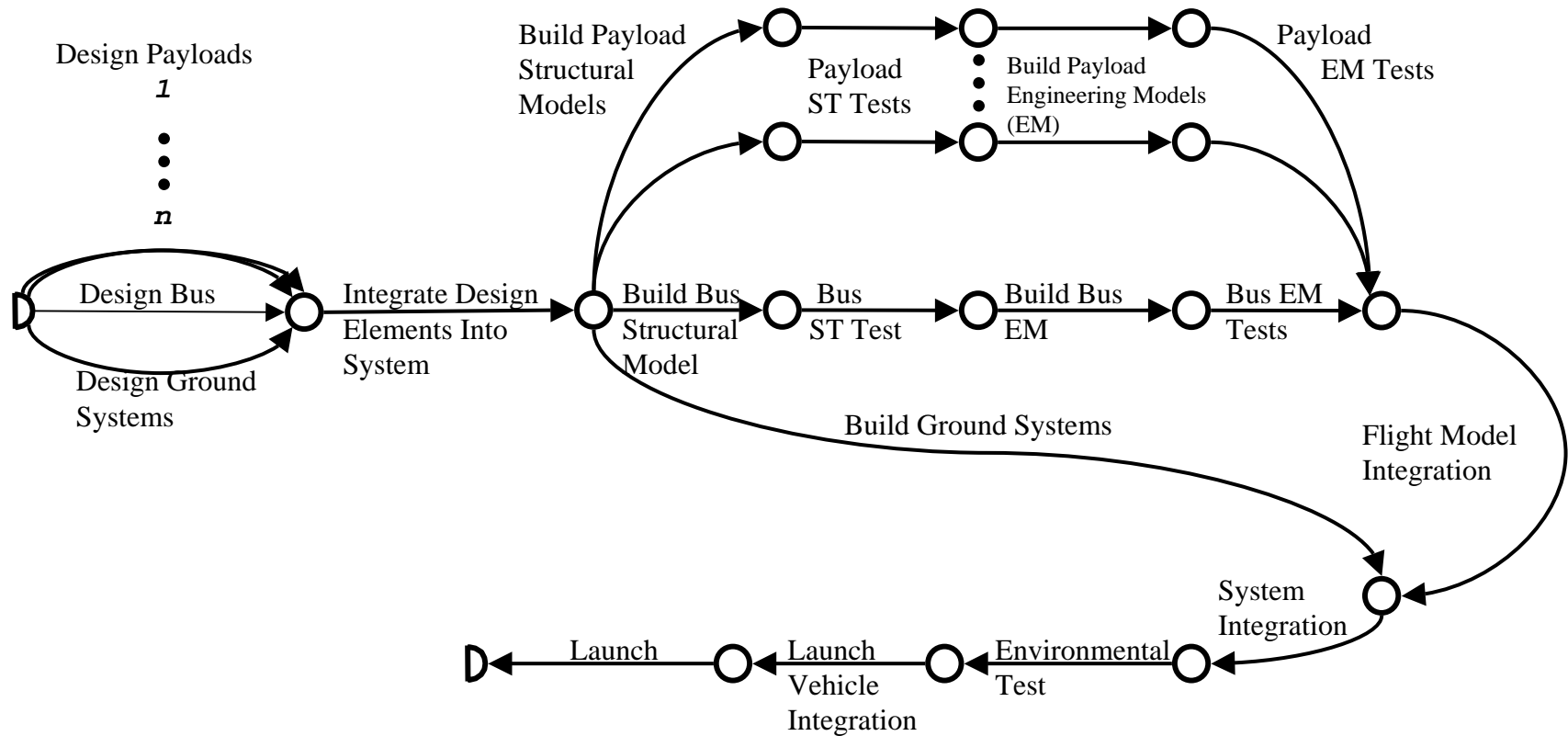


Satellite GAN Models

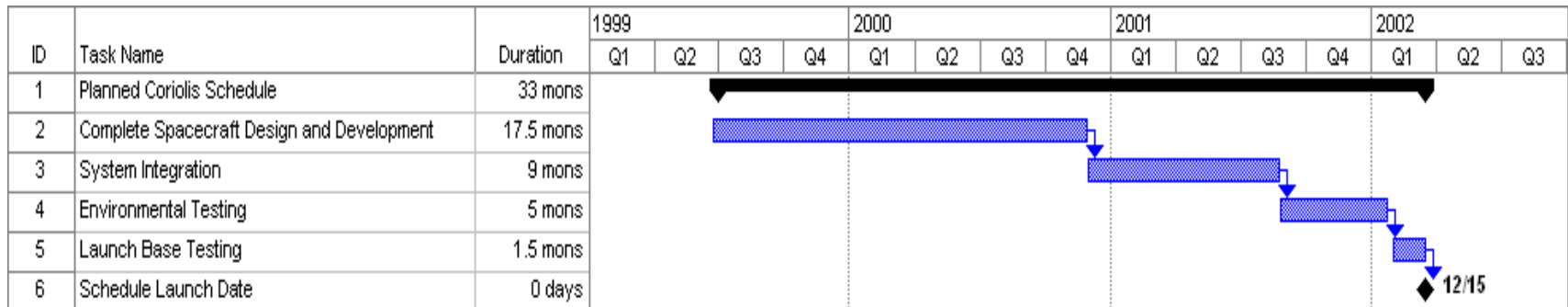
- Satellite development “good fit” for GAN modeling
- Bus and payloads almost always developed in parallel
- Satellite development process well understood
 - Initial simulations can populate baseline model
 - Refine GAN model as program develops or more data becomes available
- Extensive series of well-defined test and integration activities



Satellite Development Process



Coriolis Schedule at Contract Award



- Schedule in Fiscal Years
- Coriolis planned development schedule
 - 33-months
 - Spans Authority To Proceed (ATP) to Launch
 - Expected 12/15/01 Launch Date



Published Satellite SERs

- “Harmon, B. and Om N., (1993), “Assessing Acquisition Schedules for Unmanned Spacecraft,” *IDA Paper P-2766*, Institute for Defense Analysis, Alexandria, Virginia
- Contains satellite SERs from two Planning Research Corp. (PRC) studies: 1981 & 1990
 - PRC-D-2148 (1981): NASA data
 - PRC-D-2337-H (1990): NASA Cost Model (NASCOM) database
 - NASA and DoD satellites and unmanned space vehicles
- Burgess, Erik, (2004) “Time-Phasing Methods and Metrics” 37th Annual DoD Cost Analysis Symposium, Williamsburg, VA
 - NRO and AF Dataset



PRC Models (1981 & 1990)

- Model 1 (1981): $\text{Time in months} = 31.134 W_{FULL}^{.055}$
 $N = 21; R^2 = .02$
 - W_{Full} = launch weight of spacecraft
 - Model not statistically valid
- Model 2 (1990): $\text{Time in months} = 8.173 W_{DRY}^{.238}$
 $N = 18; R^2 = .57$
 - W_{Dry} = dry weight of spacecraft
 - Newer data set
 - Reasonable R^2 given bivariate model specification



IDA Models (1993)

- Model 3: Full data set of unmanned orbiting spacecraft

$$1stDel = .637(BOL\ Power)^{.508}(DESLIF)^{.177}1.585^{SENSOR}1.513^{NAV}.751^{COMMER}1.381^{EXPR}$$

$$R^2 = .93 \quad Adjusted\ R^2 = .90 \quad SEE = .116$$

$$Intercept\ Adjustment = 1.007 \quad N = 21$$

- BOL = Beginning of Life Power (Watts)
 - DESLIF = Design Life in Months
 - SENSOR, NAV, COMMER, and EXPR are dummy variables for whether the spacecraft are primarily sensor or scientific instrument, navigation, commercial, or experimental/scientific spacecraft
- High goodness-of-fit



IDA Models (1993)

- Model 4: Data restricted to sensor unmanned orbiting spacecraft

$$1stDEL = 2.295 \text{ BOL Power}^{.479}$$

$$R^2 = .85 \quad \text{Adjusted } R^2 = .83 \quad SEE = .156$$

$$\text{Intercept Adjustment} = 1.012 \quad N = 10$$

- BOL = Beginning of Life Power (Watts)
- Full model (Model 3) dominates sensor only model
 - Reduced goodness-of-fit
 - Fewer observations



Burgess Model (2004)

- Model 5: (2004) Time to First Launch Availability (TT1L)

$$TT1L = 17.0 + 0.87 W_{DRY}^{.406} (DESLIF * PYLD)^{.136}$$

$$Pearson's R^2 = .69 \quad SEE = .25 \quad N = 56$$

- W_{Dry} = dry weight of spacecraft
 - DESLIF = Design Life in Months
 - PYLD = Number of Payloads with Physically Distinct Hardware and Different Users
- Most recent data set; contains NRO and AF satellites

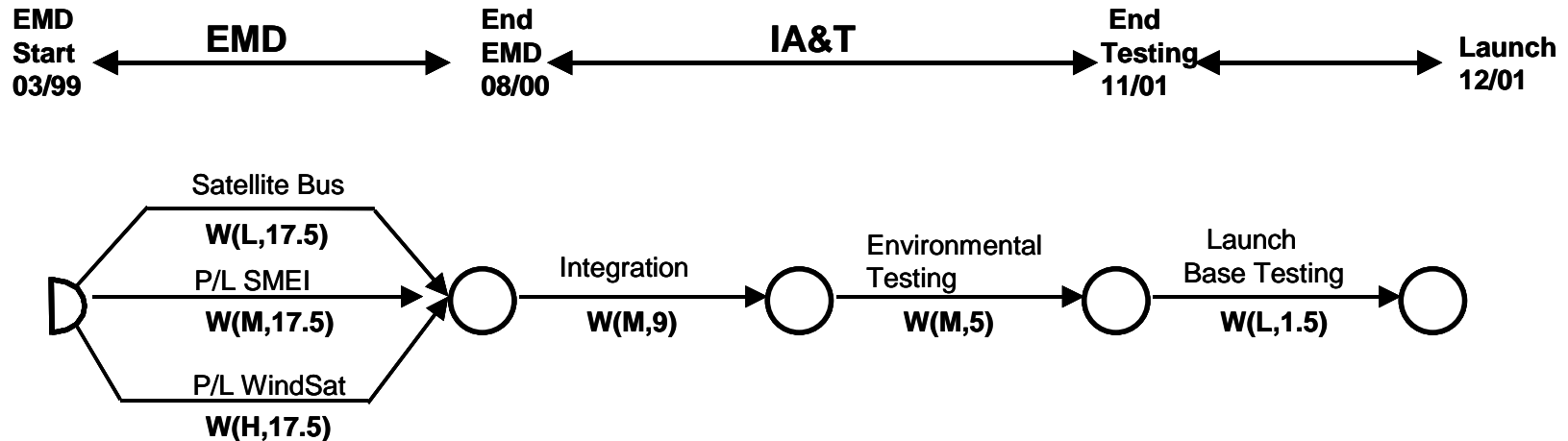


Coriolis SER Input Parameters

W_{FULL}	1801.2	<i>SENSOR</i>	1
W_{DRY}	1620.4	<i>NAV</i>	0
<i>BOL Power</i>	1209	<i>COMMER</i>	0
<i>DESLIF</i>	36	<i>EXPR</i>	0
<i>PYLD</i>	2		

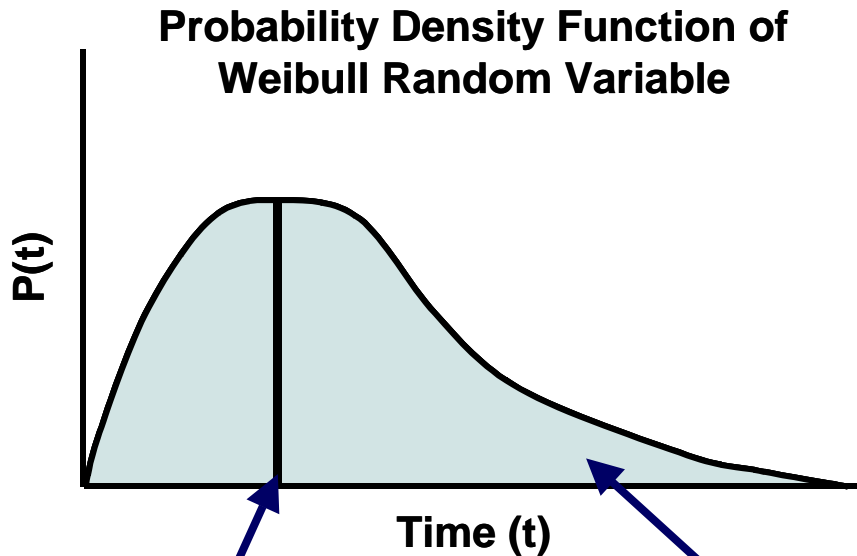


Coriolis GAN



- GAN directly constructed from initial Coriolis schedule
 - Activity arc durations modeled as Weibull distributions (Gladstone-Miller method)
 - Most likely durations are initial planned durations
 - Risk levels assigned by
 - Previous research with satellite programs (Environmental Testing, LBT)
 - Publicly available press releases dated by ATP
 - Consultation with vendors

Weibull Distributions for Time Estimates



Mode indicates most likely duration

More risk means greater likelihood that duration exceeds mode (greater skew in PDF)

- OSD PA&E has had success using these distributions to predict durations
- Can describe a Weibull distribution by specifying its mode and a *risk level*



Weibull Distributions for Time Estimates

Set Typical of More Mature Schedules

Risk	Ratio Of Mode to Min	$P(t > t_{\text{mode}})$
L	1.15	0.60
M	1.25	0.70
H	1.50	0.80

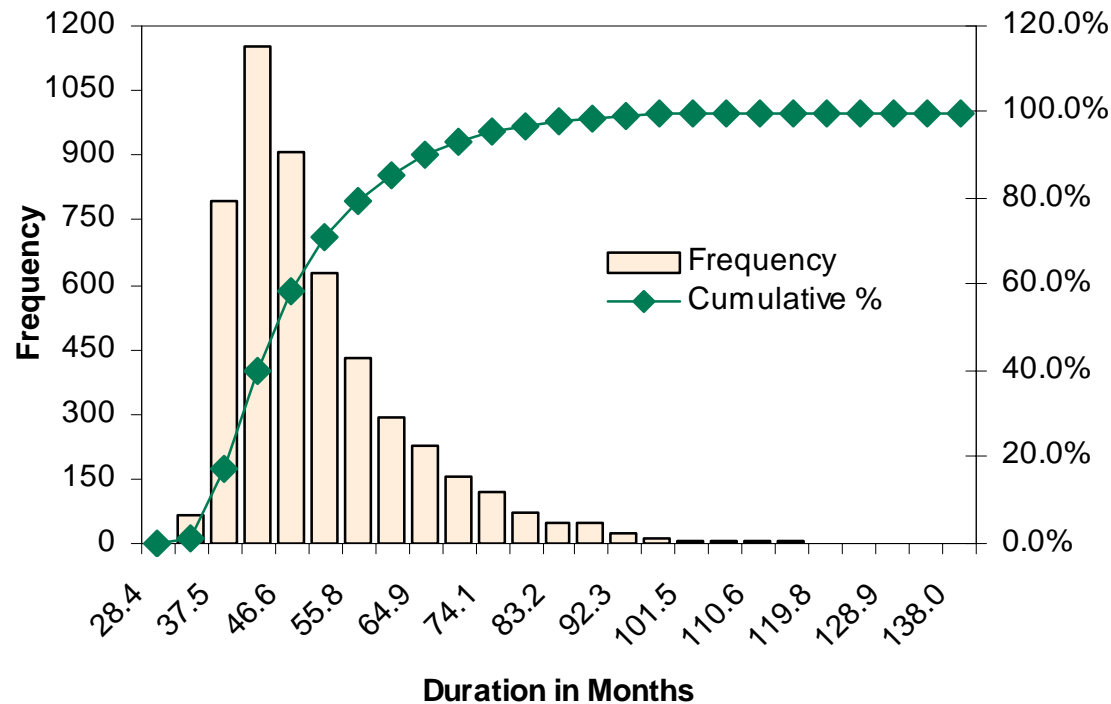
Set Typical of Less Mature Schedules

Risk	Ratio Of Mode to Min	$P(t > t_{\text{mode}})$
L	1.15	0.65
M	1.25	0.75
H	1.50	0.85

- Gladstone, B. and Miller, S. (2002), "Chemical Demilitarization Program Schedule Risk Assessment," *35th Annual DoD Cost Analysis Symposium*, Williamsburg, Virginia



Simulation Results of Coriolis GAN



- Simulation produces full distribution of outcomes
- Can identify likely duration at any specified confidence level
 - Typically use Expected Value, 50% and 80%



Findings

Model	Estimated Duration (Mths)	Estimated Schedule Slip (Mths)	Estimated Schedule Slip (%)	Actual Estimation Error (%)
Model 1	47.0	14.0	42.4%	5.6%
Model 2	47.8	14.8	44.8%	7.4%
Model 3	70.1	37.1	112.4%	57.5%
Model 4	68.8	35.8	108.5%	54.6%
Model 5	48.3	15.3	46.4%	8.5%
GAN Model				
<i>Mean</i>	47.9	14.9	45.2%	7.6%
<i>50% CDF</i>	44.5	11.5	34.8%	0.0%
<i>80% CDF</i>	56.1	23.1	70.0%	26.1%

	Duration	Slip (Mths)	Slip (%)
<i>Actual</i>	44.5	11.5	34.8%

- Must discount Model 1 results as chance
- GAN and Models 2 and 5 perform well relative to actual duration; all model predictions are upward biased



Findings

- GAN & SER Models 2 and 5 perform very well
- SER Models 3 & 4 perform poorly
 - Possible explanation?
 - Coriolis not representative of other satellites in sample
 - Different acquisition strategy
 - Micro-satellite development
 - Mature COTS bus
 - Separate, wholly independent payloads
- GAN models and SERs are good complements
 - Serve as cross-checks
 - Can use SERs early when program is ill-defined
 - As program is better understood, GANs can provide additional insight and information beyond estimate



Agenda

- Background
- What is a Generalized Activity Network (GAN)?
- Advantages & Disadvantages to GAN Approach
- GAN Performance vs. SERs: Coriolis Case Study
- Conclusion



Conclusion

- GAN models increasing in popularity among cost analysts
 - Powerful, easy to model simulations for estimation
 - Explicitly models risk and uncertainty
 - Provides additional managerial information as program evolves and GAN model is refined
- Case study indicates that GAN models perform at least as well as traditional regression-based methods
 - Further research for different commodities and test-block GANs
 - Need to investigate objective, data-driven calibrations for cost/duration distributions



LMI Authors

Danny R. Hughes
LMI & University of South Alabama
251-460-6194
dhughes@lmi.org

York W. Whitaker
Research Fellow
703-917-7045
ywhitaker@lmi.org

David Lee, Ph.D.
Senior Research Fellow
703-917-7557
dlee@lmi.org

Jeremy M. Eckhause
Research Fellow
571-633-7726
jeckhause@lmi.org



Back-up Slides

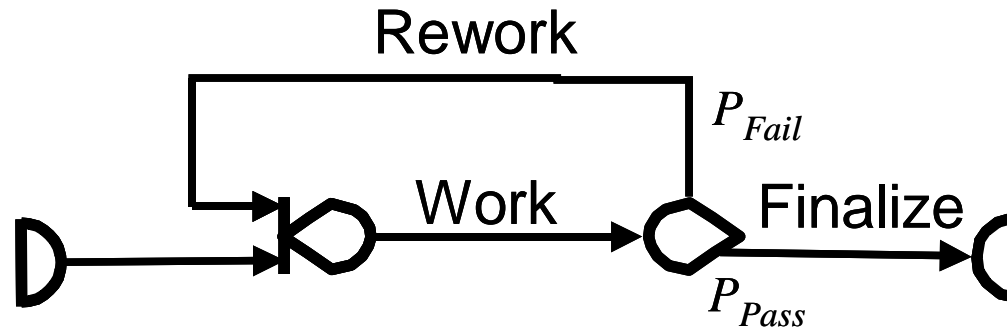


Calibrating GAN Probabilities

- We consider two common GAN feedback processes
 - The One “P” Case
 - Single feedback loop with a constant probability of success
 - Preliminary results included in MORS presentation
 - The Two “P” Case
 - Successive attempts after the first failure possess a constant, but higher, probability of success than the first test trial
 - Presumes that most of the major problems are at least identified after recovery from initial failure implying a higher probability of success for subsequent trials



GAN Probabilities: One “P” Case



- Typically, probabilities of success or failure driven by expert opinion
- Probabilities *can* be appropriately calibrated by historical data
- Assumptions
 - Well defined, common test event for commodity/system
 - Access to historical data from similar systems

GAN Probabilities: One “P” Case

- Considering simple test-block GAN:
 - Trials occur until a success is achieved (with probability P for each trial)
 - Let X be the number of trials until the first success
 - X is a geometric random variable with parameter P
 - Specifically,
$$E[X] = \frac{1}{p}$$
- Assuming historical data (of sample size n) on number of trials from similar systems can solve for single p^* that minimizes the sum of squared errors between the expected number of trials predicted by the GAN, $E[X]$, and the historical data



GAN Probabilities: One “P” Case

- Thus, if $x = \frac{1}{p^*}$ and $\{b_1, b_2, b_3, \dots, b_n\}$ are the set of outcomes representing the number of trials for independent outcomes of the same GAN, we wish to:

$$\min \sum_{i=1}^n (x - b_i)^2$$

subject to : $x \geq 1$

- Conveniently, the global minimum is simply the mean of the historical data, yielding:

$$p^* = \left(\frac{\sum_{i=1}^n b_i}{n} \right)^{-1}$$



One “P” Case: Proof

- Since our problem is only over one dimension, we can simply consider looking at the derivative of the function with respect to x

$$\sum_{i=1}^n (x - b_i)^2 = \sum_{i=1}^n (x^2 - 2b_i x + b_i^2) = nx^2 - 2x \sum_{i=1}^n b_i + \sum_{i=1}^n b_i^2$$

- Taking the derivative of this expression and setting it to zero, we get that:

$$2nx - 2 \sum_{i=1}^n b_i = 0 \quad \Rightarrow \quad x = \frac{\sum_{i=1}^n b_i}{n}$$

- Thus, we can estimate p^* by simply by taking the inverse of the average of the outcomes of the trials

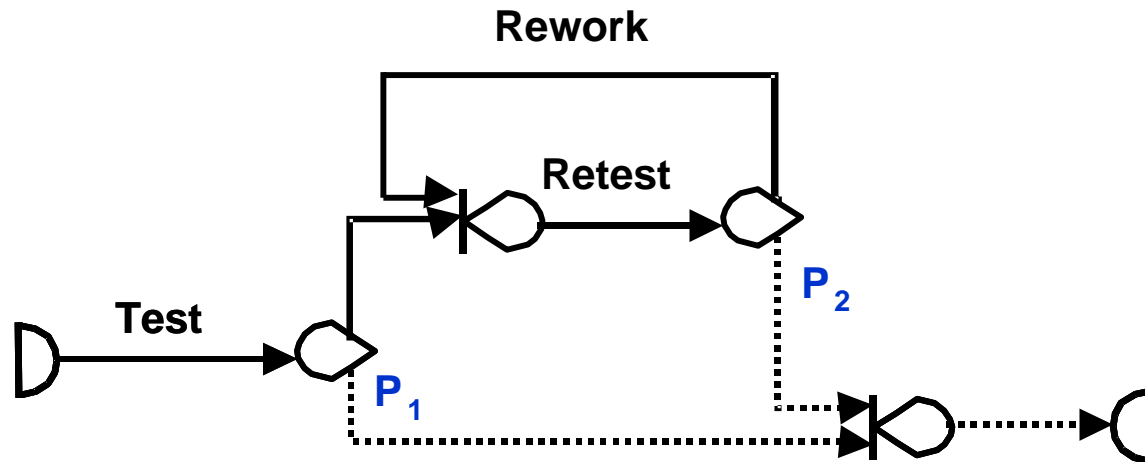


GAN Probabilities: One “P” Case

- This simple, straightforward result is powerful because analysts can easily **objectively** calibrate GAN probabilities
- Further, in absence of historical data, analysts should seek
 - Unbiased expert opinion on “average” number of tests until success
 - Should produce better estimates of realistic probability of success than directly asking for them



GAN Probabilities: Two “P” Case



- Probability of success on first test: P_1
- Probability of success on every other test, *conditional* on first test failing: P_2
 - Might expect $P_2 > P_1$ due to knowledge of what failed, additional effort spent on that item, etc.

GAN Probabilities: Two “P” Case

- Consider a test event with the following historical data:

Historical Program	# Trials until Success	1st Trial Success? (Yes=1, No=0)	2nd Trial? (Did the 1st trial fail?)	# of "P2" Trials
1	6	0	1	5
2	7	0	1	6
3	4	0	1	3
4	1	1	0	0
5	8	0	1	7
6	1	1	0	0
7	2	0	1	1
8	1	1	0	0
9	12	0	1	11
10	4	0	1	3

- We could calculate a single probability, p , using the previous technique
 - Method of calibrating P_1 and P_2 should reduce to One “P” case if probabilities are constant



GAN Probabilities: Two “P” Case

- Let x_1 and x_2 be decision variables and $\{b_1, b_2, b_3, \dots, b_n\}$ historical data.
- Let $b_1^i = 0$ if the first trial failed and $b_1^i = 1$ if it succeeded and assume that there are J successes.
- Let b_2^j represent the number of subsequent trials with a probability, p_2 , of success
- As before, we wish to minimize the sum of squared errors between the expected number of trials predicted by the GAN and the historical data for each decision node:

$$\sum_i (x_1 - b_1^i)^2 + \sum_j (x_2 - b_2^j)^2$$



GAN Probabilities: Two “P” Case

- We can minimize each sum separately, yielding x_1 and x_2 , and thus our P_1 and P_2
- Using the data from our example we produce the probabilities:

$$P_1 = x_1 = 0.3$$

$$P_2 = \frac{1}{x_2} = \frac{1}{\frac{36}{7}} = 0.1944$$

- Monte Carlo testing demonstrates method to provide robust estimation of data generating process even when $P_1 = P_2$



GAN Probabilities: Small Data Samples

- Calibrated probabilities are sensitive to sample size of available data
 - Acquisition data rarely possess sufficient sample size to appeal to asymptotic properties
- Currently examining the appropriate calculation of confidence intervals for calibrated probabilities
 - Monte Carlo testing with different sample sizes
 - Analytical derivation of confidence intervals of P_1 and P_2

